High-speed SRAM Cache Project Proposal

Lauren Cash <lec2yy>
Chuhong Duan<cd8dz>
Rebecca Reed <rcr4dd>
Andrew Tyler <adt2bt>
University of Virginia
ECE 4332 Los tOHMales Call$_e$ntes

ABSTRACT

This paper serves as a proposal for our semester project: the design and simulation of a 64 Kb embedded high-speed cache. We state which design applications we intend to target, provide simulations figures to show our preliminary functioning design drafts, describe the architectural approaches we will make and finally provide initial ideas on how to optimize our current design.

1. TARGETED DESIGN APPLICATIONS

Ideally, we will be able to design and optimize both a low power memory and a high-speed cache, in the scope of this class. However, to mitigate the risk that we only have two partially-complete designs, we will focus first on creating the 64 Kb cache and optimizing its functionality. An important research effort for us is to minimize the time it takes for the cache to read/write to cells. After creating a functional and optimized cache, we intend to modify the sizes and architecture of our design to build a 1 Mb SRAM with low power consumption as the goal. We recognize that, logically, these designs will be very similar. However, in order to accomplish both of our goals, it makes sense to attack one design and complete it fully before modifying it to create the other. We would much rather finish the project with one complete design and another one in the works (or functional, but not fully optimized) than have two close-to-completion designs.

2. PROPOSED ARCHITECTURE

Our current architecture is one that implements the most basic SRAM building blocks. Specifically, our design is drafted in a square architecture with the following dimensions: 64x64 bit cells per block, 16 blocks per array to create a 64 Kb architecture. We use basic decoders made up of a collection of AND gates. We employ a differential voltage sense amp modeled off of Rabaey's design in our textbook *(Rabaey, Digital Integrated Circuits, 2003)*. While this is a functional design, we recognize the inefficiencies that are associated with our current architecture and plan to take various steps to optimize the design. Our ultimate focus is to create a fast, streamlined cache.

The first step is to revise our decoder design. By using a conglomeration of AND gates, we decrease the speed of our cache while increasing the net area of our architecture. Neither of these is desirable. We are currently researching various techniques to improve the design of the decoder (please see the next section for details). Next, we will look at the sense amp design. Our first draft of the architecture used a design from our textbook, but we believe we can find faster, more efficient designs as we proceed with the project and our preliminary research is cited below. Finally, we plan to make executive design decisions about the shape of the cache. Our current design is formatted as a square block. Since we are choosing to optimize the speed of our design, we will make nominal sacrifices in power consumption and the total area to ensure that we have the fastest cache possible.

As the project develops, we plan to simulate many variations of our architecture in order to determine which design is the fastest. This process will be evolutionary and will provide a thorough evaluation of designs and will guarantee that our team arrives at the most optimized solution.

3. SPECIAL FEATURES

In order to optimize our design for the highest performance possible, we are looking into optimized topologies for the peripheral circuitry - decoders, sense amps, and precharge operation. The following patents will be important to test and improve upon in order to ensure we are implementing high-speed designs.

Decoders optimized for speed

- US Patent No. 5604712 - Fast Word Line Decoder for Memory Devices
- US Patent No. 7143257 - Method and Apparatus of a Smart Decoding Scheme for Fast Synchronous Read in a Memory System

Sense amps optimized for speed

- US Patent No. 5991217 - Fast SRAM Design Using Embedded Sense Amps
- US Patent No. 5698998 - Fast, Low Power, Differential Sense Amplifier

Precharge operation optimized for speed

- US Patent No. 7499312 - Fast, Stable, SRAM Cell using 7 Devices and Hierarchical Bit/Sense Line

Looking at these patents will give us a good place to start simulating different topologies and finding the optimal combination for highest performance level.

4. TIMELINE

Throughout the semester, we plan to have weekly group meetings (Tuesdays after class) and regular meetings with the project advisor (Professor Stan, by appointment) to make concrete progresses on this project. See below for our planned schedule and intermediate milestones.

Table 1. Proposed timeline for Project

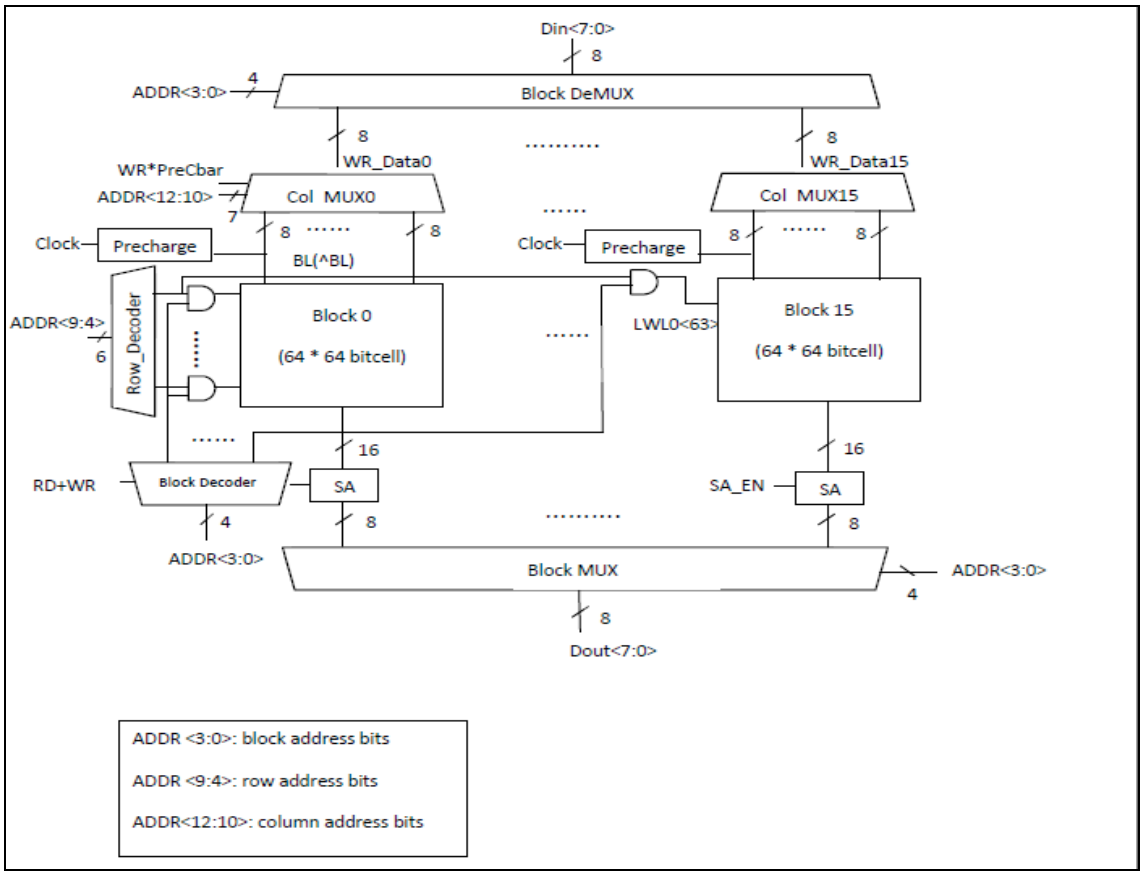
Dates	Completion
Oct 23	Complete SKILL Tutorial
Oct 30	Functioning SRAM; Optimization methods
Nov 13	Optimizations decided and implemented
Nov 18	Paper draft & compile
Nov 27	Final design & paper
Dec 2	Presentation practice

5. TASK BREAKDOWN

In order to help all group members learn key memory design principles and simulation/layout design skills we will follow an agile development process where we make assignments for group members as jobs and tasks come up. We all plan on being involved in each step and critiquing each other as the project progresses.

6. SIMULATION FIGURES

Figure 1. Our Team's Revised, Un-optimized Block Diagram



Our current design is 8*64 words per block (8 bit words - 64*64 bitcells per block) and 16 blocks in the array. The team's intuition is that a square structured memory block should have the best performance in terms of delay. More research and discussion will be addressed for structure optimizations.

Figure 2. Simulation results from Design Review 1, Simulation at Regular Conditions

Figure 3. Simulation at Process Corner (V = 1.2V, T = 0o C)

The group's simulations show that the SRAM is working functionally. As pointed out by the project advisor on the feedback on Design Review, for the simulation set-up, the group should have the Precharge signal active high, and not charging when writing. Moreover, the word line should be the last one to be asserted and the first one to be de-asserted, all the other signals should be stable during that period. The design team will work on the revisions next while making optimizations of the current design according to the metrics.

Group Wiki Page - http://venividiwiki.ee.virginia.edu/mediawiki/index.php/ClassECE4332Fall12Group_Los_Tohmales_Calientes